

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Természettudományi Kar

Földtudományi Doktori Iskola



Big Data és térbeliség

Nagy méretű raszteres téradatok elosztott számítási környezetben
történő feldolgozásának elméleti és módszertani kutatása

Doktori disszertáció tézisei

Olasz Angéla

Témavezető:

Dr. Gede Mátyás PhD adjunktus

Térképészet-Geoinformatika doktori program
Prof. Zentai László, DSc egyetemi tanár

Földtudományi Doktori Iskola
Prof. Nemes-Nagy József, DSc egyetemi tanár

Budapest

2018

Bevezetés

Napjainkban a téradatok korábban elképzelhetetlen mennyiségben kerülnek előállításra. Emellett a nyíltadat-szemlélet előretörésének és az adatpolitikai nyitásnak köszönhetően is egyre több, nagy mennyiségű, változatos téradat érhető el (pl.: ESA Copernicus program, NASA Landsat program stb.), amely indukálja a téradatokra épülő új szolgáltatások kialakítását és új generációs szolgáltatások fejlesztését (Kemp 2015, EIJ 2016, Coumans 2017). Törvényszerű volt, hogy az informatikában már széles körben használt új a nagy tömegű adatok tárolására és feldolgozására alkalmas megoldások (Big Data technológiák) a téradatok kezelésében is egyre nagyobb teret hódítsanak. A távérzékelés és térinformatika tárgyú kutatás-fejlesztés egyik legfontosabb feladata a nagy tömegű adatok feldolgozására alkalmas megoldások adaptálása, hiszen a jelenlegi tárolási és feldolgozási képességeink korlátozottak, nem képesek felvenni a versenyt az előálló adatok tárolása, feldolgozása és közzététele tekintetében.

Kutatásomban az említett Big Data technológiák fejlődését, megismerését és térinformatikai felhasználásának elméleti és módszertani kutatását tűztem ki célul. A feldolgozási eljárások fókuszában a raszteres téradatok elosztott feldolgozásának vizsgálata áll. Munkám során a nemzetközi és hazai irodalomban megtalálható definíciók áttekintését követően javaslatot teszek a Térbeli Big Data (magyar nyelvű) integrált definíciójára. Megfogalmazom a Térbeli Big Data feldolgozására alkalmas rendszer elvárásait, a raszteres adatforrásokra fókuszálva. Áttekintem az alkalmazott infrastrukturális és számítási környezeteket, különös figyelmet szentelek a Térbeli Big Data raszteres adatokat feldolgozó keretrendszerekre. Ismertetem a hagyományos Big Data és a Térbeli Big Data feldolgozási lépéseinek lehetséges különbségeit. Megfogalmazom a „Térbeli Big Analitika” leggyakrabban használt feldolgozási eljárásainak elosztott feldolgozó környezetbe történő adaptálási elvárásait, valamint számba veszem a jelenleg elérhető külföldi és hazai raszteres Térbeli Big Data platformok tulajdonságait.

Célkitűzések

1, A Big Data koncepció eredete, annak fejlődése és a napjainkban használt **Big Data** kifejezés tartalmilag nagy változáson ment át a 1997-es első (azonos értelmezésű) megjelenése óta. Azóta a kifejezés széles körben elterjedt, azonban annak teljes jelentéstartalmát kevesen ismerik, ezért szükséges annak **komplex áttekintése és ismertetése**. A *Térbeli Big Data (Geospatial Big Data, Spatial Big Data* stb.) kifejezés az elmúlt 5 évben jelent meg (azonos elnevezéssel és jelentéstartalommal), de annak jelentése a korai meghatározásokban kizárólag a nagy mennyiségű téradatra korlátozódott. Szükséges egy **átfogó definíció megalkotása** a „Térbeli Big Data”-jelenség komplex leírására, ami magába foglalja a rendelkezésre álló téradatok mennyiségi és egyéb tulajdonságait, a szükséges feldolgozó környezetet és a megfelelő elemzési eljárásokat is.

2, A **Térbeli Big Data adatforrásainak azonosítása és tulajdonságainak áttekintése**, kiemelten fontos a hatékony elosztott feldolgozás érdekében. Egy Térbeli Big Data platform kiépítésének első lépése a feldolgozandó téradatok azonosítása és komplex jellemzése.

3, Az elosztott tárolás és feldolgozás alapja az arra alkalmas **infrastruktúra és számítási környezet**, ezért szükséges az ezt biztosító architektúra típusok, felügyeleti eszközök és tároló- és feldolgozó **keretrendszerek áttekintése és jellemzők ismertetése**. Ennek elérése érdekében a rendelkezésre álló megoldásokat példákon keresztül tanulmányozom, majd elvégzem a szolgáltatásaik, funkcionalitásuk és képességeik összehasonlítását egy **általánosan felállított szempontrendszer szerint**, amely elsősorban a raszteres adatok elosztott feldolgozásának alapvető igényeit összegzi. Gyakorlati célom mindezek alapján az **infrastruktúrával és számítási környezettel szembeni elvárások megfogalmazása** a raszteres téradatok elosztott feldolgozásának megvalósítása érdekében.

4, A **Big Analitika** fogalmának meghatározása után szükséges az **eszközrendszerének tanulmányozása** és a hagyományos Big Analitika és a Térbeli

Big Analitika feldolgozási folyamatának lehetséges eltéréseinek azonosítása. Célom ezen eltérések feltárása és a raszteres adatfeldolgozás során szükséges **kiegészítő feldolgozási lépések meghatározása**. A fentiek alapján további célom az **analitikai környezet elvárásainak meghatározása** a raszteres Térbeli Big Data feldolgozásában. Ezen túlmenően szükséges az elosztott feldolgozás előnyeinek kihasználása érdekében a **képfeldolgozó algoritmusok párhuzamosíthatóságának vizsgálata** és az erre szolgáló eljárások rendszerbe foglalt áttekintése.

5, A Térbeli Big Data aktuális helyzetének meghatározására jó eszköz, a gyakorlati felhasználás megalapozásként a **jelenleg létező komplex rendszereket, platformokat áttekintő jellegű ismertetése**, konkrét **nemzetközi és hazai példák**on keresztül. Célom az áttekintés strukturált elvégzése, mely kiterjed a támogatott fejlesztői környezet ismertetésére.

Alkalmazott módszerek

A kutatásom során végigkövettem a Big Data¹ definíciójának fejlődési útját a releváns szakirodalom alapján, majd összegeztem annak főbb elemeit. Továbbá áttekintettem a Térbeli Big Data definícióit a vonatkozó legfontosabb szakirodalmak alapján. A Térbeli Big Data fogalmának nagy értelmezési különbségei miatt szükségesnek tartom komplex definíciójának megfogalmazását.

A Térbeli Big Data adatforrásainak azonosításakor a különböző szakirodalmi források (szakkönyvek, lektorált szakfolyóiratok, szakterület-specifikus online folyóiratok, interjúk, különböző műhely publikációk, projektjelentések stb.) szintetizálását végeztem el. Az adatforrások csoportosítását és jellemzését saját szakterület specifikus tapasztalataim hozzáadásával készítettem el.

#####

¹A dolgozatban egységesen használom a Big Data írásmódot nagy kezdőbetűvel. Minden esetben az új technológia egészét jelenti, nem az adatmennyiségi értelmezéseket. (Hasonló megfontolásokat találtam az ENSZ Big Data tananyagában is, UN SSC 2018). A Térbeli Big Data nagy kezdőbetűs írását is ezért tartom meg.

A Térbeli Big Data keretrendszerek összehasonlítását egy általam felállított szempontrendszer alapján végeztem. A szempontrendszer az infrastruktúra, a tároló- és feldolgozó keretrendszer, valamint az analitika elvárásaihoz igazodik.

A Térbeli Big Analitika feldolgozási, elemzési eljárásainak gyűjteménye a legelterjedtebb eszközöket tartalmazza. A gyűjtemény összeállításában az IQmulus projekt tapasztalatait is figyelembe vettem. Az adatparticionálási és -összeillesztési metódus szintjeinek általános javaslatát az IQmulus és IQLib eredményeinek továbbfejlesztésével készítettem el.

A külföldi rászteres Térbeli Big Data platformok példáinak kiválasztásában az általam tanulmányozott infrastruktúra és keretrendszerek felhasználásának megalapozottságát tartottam szem előtt. A kiválasztásban a nyílt elérés, a nemzeti szintű földmegfigyelési feladatok ellátása, vagy a már korábban létrehozott platformok szolgáltatásainak összekapcsolása kiemelt szempontjaim voltak. A hazai Térbeli Big Data platformok esetében nem csak a rászteres feldolgozásra koncentráltam, hanem a téradatok új típusú felhasználására is hoztam példát.

Eredmények

1. A konzisztens definíció megalkotása érdekében áttekintettem a rendelkezésre álló Térbeli Big Data definíciókat, összefoglaltam azok tartalmát, majd **meghatároztam egy átfogó definíciót a „Térbeli Big Data”-jelenség komplex leírására, amely magában foglalja a rendelkezésre álló téradatok mennyiségi és egyéb tulajdonságait, a szükséges feldolgozó környezetet és a megfelelő elemzési eljárásokat is.**

A korai Big Data definíciókhoz hasonlóan a jelenleg elérhető Térbeli Big Data meghatározások csak részben érintik az egyes elemeket, csak a nagy térbeli adatok jellemzőire fókuszálnak, esetleg kiegészítik, hogy eltérő nagy kapacitású feldolgozó, tároló környezetet igényelnek. Az elérhető Térbeli Big Data (*Geospatial Big Data*, *Spatial Big Data*) fogalmainak áttekintése alapján

kijelentem, hogy nincs átfogó, komplex megfogalmazás a tér adatok (térbeli és időbeli felbontása) tulajdonságainak eltérő kezelési igényeit kielégítő Big Data technológiát felhasználó meghatározásra példa. Hiánypótló annak meghatározása a témával foglalkozó kutatások kiindulásaként.

<i>A Térbeli Big Data technológia nagy mennyiségű, nagy sebességű és igen változatos tér adat (vektor, raster, pontfelhő, helycímkézett szöveges adatok) gyors és hatékony feldolgozására képes, ehhez felhasználja az innovatív Térbeli Big Analitika eszközeit (térinformatikai, rasteres képfeldolgozó matematikai-statisztikai algoritmusok összessége) és a modern Big Data infrastrukturális és számítási környezet elemeit, így lehetővé válik a tér adatok összekapcsolt feldolgozása, adatcseréje, és megjelenítése. Ezáltal megvalósítja a nagy adatokon alapuló térbeli és időbeli döntéstámogatás technológiai alapját, amely különös figyelmet szentel az tér adatok forrásainak megbízhatóságára és a biztonsági elvárásokra.</i>	Olasz A. 2018
---	------------------

Az 1. tézis témájában megjelent publikációm: Olasz A. (2018).

2. Azonosítottam, csoportosítottam és jellemeztem a Térbeli Big Data adatforrásait a *mennyiség, változatosság, sebesség és megjelenítés* dimenziók szerint, különböző adattárolási típusokra. Összehasonlítottam a hagyományos „Big Data”, a „Térbeli Big Data” és a „tradicionális” tér adatokat egy általam felépített szempontrendszer szerint.

A Térbeli Big Data adatforrásainak azonosítása és tulajdonságainak áttekintése kiemelten fontos a hatékony elosztott feldolgozás érdekében. A Térbeli Big Data platformok célja változatos lehet, igazodva az adatforrások változatosságához. A Térbeli Big Data adatforrásainak csoportosítása a gépek és emberek által előállított főbb csoportokra oszthatók. Ezen főbb csoportok adatforrásait csak akkor tekinthetjük a Térbeli Big Data adatforrásainak, ha azok megfelelnek a Big Data jellemzésében bevezetett dimenzióknak, valamint rendelkeznek térbeli (spektrális és időbeli) jellemzőkkel is. A jellemzés segít eligazodni a térbeli nagy

adatcsoportok, azok gyors reagálású alkalmazási területeken való felhasználása, webes, interaktív megjelenítése és az elemzésükre gyakran felhasznált eljárások között. A „Földmegfigyelési Big Data” (Guo 2017), mint a Térbeli Big Data adatforrásainak egyik alcsoportja kiemelkedő jelentőségű célterülete a fejlesztéseknek és egyben a kutatásomnak is, ezért annak eltérő dimenzióit kiemelten jellemeztem.

A 2. tézis témájában megjelent publikációim: Olasz A. (2017), Olasz A., B. Nguyen Thai és D. Kristóf (2017).

3. Az elosztott térbeli számítás szakirodalmi áttekintése után a téradatok elosztott feldolgozására alkalmas infrastruktúrával és a számítási környezettel szemben támasztott magas szintű általános és szakterület-specifikus követelményeket fogalmaztam meg. Az általam kialakított szempontrendszer szerint jellemeztem a jelenleg elérhető legjelentősebb nyílt forráskódú feldolgozó keretrendszereket.

Könnyen belátható, hogy egyre nagyobb számítási teljesítményt kell elérni ahhoz, hogy a Big Data feldolgozás elvégezhető legyen, ezen kívül szükség van olyan eljárásokra, amelyek segítségével képesek vagyunk az adatok szűrésére, redukálására az értékes információ kinyerése érdekében. Az elemzéseket belátható időn belül csak elosztott környezetben lehet elvégezni, ahol az egyes munkafolyamatok képesek párhuzamosan futni az elosztott adatokon. A Térbeli Big Data feldolgozásának eszközeit adja a Big Data technológia által kínált hardver- és szoftver-architektúra, de azok felhasználásának körülményei nem triviálisak. Azonban a térinformatikai, távérzékelési (beleértve a fotogrammetriát is) eljárások elvégzését a hagyományos Big Data adatforrásokra kidolgozott rendszerekkel elvégezni nem lehetséges. Általánosságban elmondható, hogy a Big Data technológiák alkalmazása a téradatok feldolgozásakor nagyfokú körültekintést igényel az optimalizált, gyors, szakterület-specifikus eredmények elérése érdekében.

Általános elvárások:

1. skálázhatóság biztosítása: a rendszer áteresztőképességének a rugalmas növelésére legyen lehetőség újabb szerverek, komponensek (CPU, RAM, HD) hozzáadásával.
2. hibatűrés: a működés során bekövetkező hibák észlelése, javítása és kiküszöbölése érdekében álljon rendelkezésre felügyeleti eszközpark. A hibatűrésnek több fokozata létezik, amelyek a műveleti eredmények ellenőrzésével, az információk több példányban történő tárolásával, illetve redundáns hardverelemek alkalmazásával, valamint ezek kombinálásával valósíthatók meg.
3. elosztási átlátszóság (*distribution transparency*) lehetősége: ami a felhasználó elől elrejt a mögöttes infrastruktúra elemeit, legyen lehetőség rugalmasan alakítani a felhasználó profiljához igazodva (részletesebben az analitikai elvárásoknál kifejtve).
4. platformfüggetlenség: fontos, hogy a rendszer rugalmassága érdekében többféle infrastruktúrátípust támogasson, könnyű legyen migrálni egyik környezetből a másikba.
5. munkafolyamatok felügyeletének lehetősége: a munkafolyamatok indítása, megállítása, és folyamatos nyomon követése (várható feladat elvégzésére szánt idő becslése) elengedhetetlen.
6. könnyű kezelés: könnyű erőforrás-kezelési lehetőség, felhasználóbarát használat, rugalmas keresési, feldolgozási, tárolási, megjelenítési funkciók támogatása.
7. függetlenség a különböző (pl. MapReduce) programozási modelttől: az eljárás (az alkalmazott architektúrán) kisebb értelmetlen egyforma limitált méretű darabokra bontja fel a feldolgozandó adatot, amit eloszt a feldolgozó node-ok valamelyikére és végül a feldolgozott darabokat egyesíti. A rendszer támogassa az egyes csomagok felosztásának módosítására és elosztásának felügyeletét a megfelelő (szakértői) felhasználók számára. Valamint a MapReduce technológia az iteratív munkafolyamatok esetében limitált.

Az általános követelmények mellett fontosnak tartom összefoglalni azokat a térbeli eset-specifikus elvárásokat, amelyek egy térbeli adatokat feldolgozó rendszertől elvárhatunk. A követelményeket az IQmulus-projekt kapcsán szerzett tapasztalatok alapján fogalmazom meg, amelyben egy a téradatokat elosztottan feldolgozó rendszer építésében, felhasználói elvárásainak azonosításában és a térbeli feldolgozó munkafolyamatok megfogalmazásában vettem részt.

Szakterület-specifikus elvárások

- 1) a meglévő programok futtathatósága különböző eseteinek támogatása (esetfüggő).

Szintjei:

- (a) A meglévő algoritmusok teljes eszköztárát újra kell implementálni, és optimalizálni az új, megváltozott környezetre adaptálva, a masszív párhuzamos feldolgozásra. Előnye, hogy az új környezetben az elemzések gyorsabban végrehajthatók, minimális írás-olvasási költség merül fel (*overhead*). Hátránya: a rendszer kiépítésekor nagy előállítási idővel (költséggel) kell számolni.
 - (b) a térinformatikai/távérzékelési algoritmusfejlesztők megszokott kódolási környezetben dolgoznak, amelynek az új feldolgozási környezetben is elérhetőnek és futtathatónak kell lennie. Előnye, hogy nem kell újraírni a meglévő eljárásokat. Hátránya, hogy várhatóan nagyobb overhaddel – ezáltal futási idővel kell – számolni. Szükség van egy közvetítő szintre az infrastruktúra és az algoritmusok között (*middleware*). A korábban egy szálon futó algoritmusok párhuzamos futtatása nem mindig lehetséges.
- 2) a térinformatikai/távérzékelési algoritmusfejlesztők által előre megírt adatparticionálási és összeillesztési metódusok (*Tiling & stitching*) elérhetőek legyenek a nem fejlesztő felhasználók számára, ajánlásokkal, a tér adatok és esetek (*use-case*) feldolgozási céljainak megfelelően.

Szintjei:

- a) a felhasználó kiválaszthatja a meglévő eljárásokból,
 - b) a fejlesztő rendel hozzá alapbeállítást az esettől függően,
 - c) a rendszer a futtatási naplófájlok alapján ajánlja fel a hatékony stratégiákat.
- 3) optimalizált legyen sokféle adattípusra (szenzor, vektor, raszter, pontfelhő): optimalizált legyen az adatok keresésében és elérésében, valamint a már feldolgozott adatokat is könnyen kereshetővé kell tenni a rendszerben (indexelés stb.).
- 4) a tér adatok (vektor, raszter, pontfelhő, egyéb) könnyen és gyorsan legyenek betölthetőek a rendszerbe (akár publikus tárházak letöltési szolgáltatásainak automatizált elérésével, akár lokális központi adatszerveken, adathordozókon tárolt adatokra van szükség).

A 3. tétel témájában megjelent publikációim: Olasz A. (2018). Olasz, A., B. Nguyen Thai and D. Kristóf (2016)

4. A „Big Data”-jelenséghez kapcsolódóan, annak kiemelt alterületeként külön tárgyaltam a **„Big Analitika”** fogalmat. Annak definiálása után **tanulmányoztam eszközszerét**, és **azonosítottam** a „hagyományos” Big Analitika és a Térbeli Big Analitika **feldolgozási folyamatának eltéréseit**. Ennek részeként **meghatároztam a raszteres adatfeldolgozás során szükséges kiegészítő feldolgozási lépéseket**. **Meghatároztam** a raszteres Térbeli Big Data feldolgozóplatformok kialakításakor felmerülő minimális **analitikai környezet funkcionális elvárásait**. Ezen túlmenően **rendszerbe foglalt áttekintést adtam** a gyakran használt **képfeldolgozó algoritmusok párhuzamosíthatóságáról** az elosztott feldolgozás előnyeinek kihasználása érdekében. A párhuzamosíthatóság vizsgálata kiterjedt az algoritmusok **párhuzamosításának és a raszteres téradatok particionálási és összegyűjtési stratégiájának meghatározására**.

A téradatok (*spatio-temporal data*) hagyományos feldolgozásának elemei a matematikai-statisztikai módszertanra épülnek. A térinformatikai, távérzékelési statisztikai és egyéb szakterület-specifikus eljárások azonban a Big Data területére érkezve kihívásokkal néznek szembe, ezért új megoldások kidolgozására van szükség, ahogyan az a hagyományos adatelemzési eljárások Big Data képes adaptációja során történik. Az alábbi térbeli feldolgozó környezet analitikai elvárásait foglalmaztam meg:

Általános elvárások

- 1) alapvető elvárás az analitika szállítása az adathoz, az adatmozgatások minimalizálása érdekében.
- 2) az algoritmusfejlesztői környezet esetfüggő követelményrendszere.
Lehetséges szintjei:
 - a) ne legyenek programozási (pl.: Python, Ruby stb.) és algoritmikuskorlátozások a keretrendszer (pl.: R, MatLab, Python stb.) használatakor.
 - b) a gyakran használt fejlesztői környezetet támogassa a rendszer, de továbbiakat ne.
 - c) kötött fejlesztői környezet legyen, ahol csak szigorú kritériumok szerint lehet algoritmusokat implementálni.
- 3) a téradatok (amelyek lehetnek feldolgozatlanok vagy más felhasználók által előállított eredmények) keresése legyen egyszerű (szűrés és térképi lehatárolás alapján stb.)

Szakterület-specifikus elvárások

- 1) támogassa a térinformatikai és távérzékelési feldolgozó könyvtárakat, nyílt forráskódú feldolgozó szoftvereket (pl.: GDAL, SNAP, OTB, stb.).
- 2) támogassa a gyakran használt térinformatikai adattípust (formátumot).
- 3) lehessen az éppen aktuális térinformatikai és képfeldolgozási algoritmushoz megválasztani a raszteres adatparticionálási és -összeillesztés metódus szintjét (*Tiling & stitching*) a számítási csomópontokon (*node*).
- 4) A térinformatikai és távérzékelési algoritmusfejlesztők által létrehozott, adaptált cél térinformatika és távérzékelési algoritmusok legyenek tárolhatóak, kereshetők egy központi algoritmustárban.
- 5) Könnyen lehessen bővíteni az algoritmustárat.
- 6) A munkafolyamatokat szabadon kombinálható feldolgozó eljárások (*service*) láncolataként lehessen előállítani, azaz az egyes akár eltérő fejlesztők által létrehozott eljárások bemeneti és kimeneti állományai, adatformátuma egymással kompatibilis legyen.

A 4. tézis témájában megjelent publikációm: Nguyen Thai, B. és Olasz, A., (2015), B. Nguyen Thai és D. Kristóf (2017).

5. A Térbeli Big Data aktuális helyzetének áttekintése után, a gyakorlati felhasználás megalapozásként áttekintettem és jellemeztem a **jelenleg létező komplex rendszereket, platformokat**, konkrét **külföldi és hazai példák**on keresztül. Az áttekintésnél kiemelt figyelmet fordítottam a rendszerekben felhasznált térbeli számítási keretrendszerek, fejlesztőkörnyezet és a támogatott térbeli elemzések feltárására.

Az 5. tézis témájában megjelent publikációm: Olasz, A., B. Nguyen Thai, R. Giachetta és D. Kristóf (2015).

Következtetések

Vizsgálataim során arra a meggyőződésre jutottam, hogy azon túlmenően, hogy az egyes kutatás-fejlesztési projektek, elemzések bizonyítják egy-egy alkalmazási példa létjogosultságát (ami erősen esetspecifikus elemeket tartalmazhat), szükség

van a Térbeli Big Data feldolgozása során szabványok kidolgozására is. Annak érdekében, hogy az elemzések megismételhetők (ellenőrizhetők) és a kiépített rendszerek közötti kapcsolat könnyedén kialakíthatók legyenek (adatcsere, adatmegjelenítés és publikálás alkalmával) általános, platformfüggetlen Térbeli Big Data referencia-architektúra meghatározása is szükséges.

Szükséges továbbá az egyes Térbeli Big Data platformok együttműködése, ennek megvalósítására azonban csak szabványos módon kialakított adatelérés, infrastrukturális és számítási környezet, elemzési munkafolyamatok (workflow), algoritmusok felhasználása mellett van mód. Tehát nem csak a platformok kiépítése és az adatcsere, adatpublikálás során van szükség szabványos megoldások kidolgozására, hanem a Térbeli Big Analitika szabványosítása is elvárás (pl. adatleírás, algoritmus, szolgáltatási szabványok stb.) A referencia-architektúra kulcskomponenseinek meghatározása az első lépés a szabványos (lehetőleg infrastruktúrafejlesztő- és gyártó cégtől független) platformok létrehozásában.

Természetesen szükség van további eljárások áttekintésére és ilyen szempontú értékelésére annak érdekében, hogy a szakterület alkalmazkodjon a kor kihívásaihoz és a jövőben is képes legyen feldolgozni a rendelkezésre álló téradatokat.

Véleményem szerint a szakterület (hasonlóan másokhoz) paradigmaváltáson megy keresztül, mely nagyban meg fogja változtatni a jövőbeli munkavégzés eszközrendszerét.

A disszertáció témájában megjelent publikációk

2018

Olasz A. (2018). Hogyan lesz térbeli a Big Data? Nagyméretű téradatok elosztott feldolgozása, *GEODÉZIA ÉS KARTOGRÁFIA*, 2018/1: pp 25-33.

2017

Olasz A. (2017). Big Data és térbeliség, *GEODÉZIA ÉS KARTOGRÁFIA*, 2017/5, pp. 12-21.

Olasz, A., B. Nguyen Thai és D. Kristóf (2017). Development of a New Framework for Distributed Processing of Geospatial Big Data, *INTERNATIONAL JOURNAL OF SPATIAL DATA INFRASTRUCTURES RESEARCH* Vol. 12, 1212: pp. 85–111.

Olasz, A., D. Kristóf, B. Nguyen Thai, M. Belényesi és R. Giachetta (2017). Processing Big Remote Sensing Data for fast flood detection in a distributed computing environment, *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING, ISPRS (2002-)*, Vol. XLII-4/W2. pp. 137-142.

2016

Olasz, A., B. Nguyen Thai és D. Kristóf (2016). A new initiative for Tiling, Stitching and Processing Geospatial Big Data in Distributed Computing Environments, *INTERNATIONAL ANNALS OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES, ISPRS*, Vol. III-4: pp.111–118.

Független idéző: 2 Függő idéző: 3 Összesen: 5

Olasz, A. és B. Nguyen Thai (2016). Geospatial Big Data processing in an open source distributed computing environment. *PEERJ PREPRINTS* e2226v1:(4) pp. 1-4.

2015

Nguyen Thai, B. és Olasz, A. (2015). Raster data partitioning for supporting distributed GIS processing, *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING. ISPRS (2002-)*, Vol. XL-3/W3: pp. 543–551.

Függő idéző: 2 Összesen: 2

Olasz, A., B. Nguyen Thai, R. Giachetta és D. Kristóf (2015). Big Geospatial Data processing in the IQmulus Cloud, In: Orosz G. T. (szerk.) *Proceedings of 10th International Symposium on Applied Informatics and Related Areas (AIS 2015)*, (ISBN: 978-615-5460-49-4) pp. 29–35.

Olasz, A., D. Kristóf, M. Belényesi, K. Bakos, Z. Kovács, B. Balázs és Sz. Szabó (2015). IQPC 2015 Track: Water detection and classification on multi-source remote sensing and terrain data, *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING. ISPRS (2002-)*, Vol. XL-3/W3. (ISBN:978-1-4799-7929-5) pp. 583-588.

Függő idéző: 1 Összesen: 1

Olasz, A., D. Kristóf és B. Nguyen Thai (2015). Big geospatial data processing, In: Boda J. (szerk.) *Konferenciakötet: Az elmélet és a gyakorlat találkozása a térinformatikában címmel, Térinformatikai Konferencia és Szakkiállítás VI.*, (ISBN:978-963-318-488-2) pp. 21–30.

Kristóf, D., R. Giachetta, A. Olasz, M. Belényesi, B.N. Thai és M. Harsányi (2015). Using big geospatial data for fast flood detection: Developments from the IQmulus project, *Proceedings of IGARSS - IEEE International Geoscience and Remote Sensing Symposium 2015*, 7/2015: 838–841.

Független idéző: 2 Összesen: 2

2014

Kristóf, D., R. Giachetta, A. Olasz és B. Nguyen Thai (2014). Big Geospatial Data Processing and Analysis Developments in the IQmulus Project, In: P. Soille, P. G. Marchetti (szerk.) *Proceedings of the Conference on Big Data from Space (ESA) 2014*, (ISBN: 978-92-79-43252-1) pp. 214–217.

Olasz, A. és B. Nguyen Thai (2014). Decision support on distributed computing environment (IQmulus), In: A. Jolma, P. Sarkola, L. Lehto (szerk.) *Proceedings of the 3rd*

Open Source Geospatial Research & Education Symposium (OGRS 2014), Aalto University publication series, SCIENCE + TECHNOLOGY (5/2014), (ISBN: 978-952-60-5707-1) pp. 107–114.

A disszertáció témájában megtartott jelentősebb konferencia előadások

2017

A.Olasz, D.Kristóf, B.Nguyen Thai, M. Belényesi, R.Giachetta (2017) Processing Big Remote Sensing Data for fast flood detection in distributed computing environment **Free and Open Source Software for Geospatial Europe (FOSS4G-E) Academic Track**. Párizs, Franciaország: 2017.07.18 -2017.06.22.

2016

Olasz A., és D. Kristóf (2016) Development of a new framework for Distributed Processing of Big Geospatial Data, **Free and Open Source Software for Geospatial (FOSS4G) World Conference, Academic Track**, Bonn, Németország, 2016. augusztus 24-26.

Olasz, A., B. Nguyen Thai és D. Kristóf (2016). A new initiative for Tiling, Stitching and Processing Geospatial Big Data in Distributed Computing Environments, **XXIII ISPRS Congress** Prága, Csehország, 2016. július 12-19.

2015

Olasz, A., B. Nguyen Thai, R. Giachetta és D. Kristóf (2016). Big Geospatial Data processing in the IQmulus Cloud, **11th International Symposium on Applied Informatics and Related Areas (AIS 2015)**, Székesfehérvár, Magyarország, 2015. november 17.

Nguyen Thai, B. és Olasz, A., (2015). Raster data partitioning for supporting distributed GIS processing, Poszter előadás, **ISPRS Geospatial Week 2015**. Montpellier, Franciaország, 2015. szeptember 28.- 2015.október 02.

2014

Olasz, A. és B. Nguyen Thai (2014). Decision support on distributed computing environment (IQmulus), **3rd Open Source Geospatial Research & Education Symposium (OGRS 2014)**, Espoo, Finnország, 2014. Június.10-13.